

**LAPORAN TUGAS BESAR KECERDASAN BUATAN  
PREDIKSI PENUTUPAN HARGA SAHAM MNC BANK MENGGUNAKAN  
ALGORITMA REGRESI**

Dosen Pengampu:  
Leni Fitriani, S.Kom, M.Kom



Disusun oleh:  
Suci Nursania – 2306046  
Moh Ramdani – 2306062

**INSTITUT TEKNOLOGI GARUT  
JURUSAN ILMU KOMPUTER  
PROGRAM STUDI TEKNIK INFORMATIKA  
TAHUN AKADEMIK 2024/2025**

## 1. PENDAHULUAN

Perkembangan dunia finansial yang sangat dinamis menuntut tersedianya sistem prediksi yang akurat untuk membantu investor dalam mengambil keputusan. Pasar saham, sebagai salah satu instrumen investasi utama, sangat dipengaruhi oleh berbagai faktor ekonomi, politik, dan psikologis yang menyebabkan harga saham mengalami fluktuasi setiap waktu. Dalam konteks ini, prediksi harga saham menjadi salah satu tantangan penting yang memerlukan pendekatan analitik canggih, salah satunya melalui penerapan algoritma *machine learning* (Purnama Sari et al., 2024).

Penerapan kecerdasan buatan dalam prediksi harga saham menawarkan pendekatan yang adaptif, akurat, dan efisien. Algoritma *regresi* menjadi salah satu metode yang sering digunakan dalam *forecasting* harga saham karena kemampuannya dalam menangani data numerik dan tren waktu (Triya et al., 2024). Dua algoritma regresi yang banyak digunakan dalam prediksi harga saham adalah *K-Nearest Neighbor (KNN)* dan *Linear Regression*. Keduanya dikenal karena implementasinya yang relatif sederhana namun efektif. KNN melakukan prediksi berdasarkan kedekatan nilai dengan data historis, sedangkan Linear Regression mencoba membangun hubungan linier antara variabel independen dan dependen. Berbagai studi sebelumnya menunjukkan bahwa kedua algoritma ini memiliki performa yang cukup baik dalam konteks prediksi harga saham, tergantung pada karakteristik datanya (Eko Waluyo et al., n.d.).

Seiring dengan meningkatnya akses terhadap data historis pasar saham dan berkembangnya perangkat analitik berbasis AI, pemanfaatan metode *machine learning* dalam prediksi pasar saham tidak lagi bersifat eksperimental, melainkan menjadi kebutuhan yang nyata. Sistem prediksi berbasis data mampu meningkatkan efisiensi analisis teknikal, menyediakan wawasan kuantitatif yang objektif, serta mengurangi ketergantungan pada intuisi subjektif investor. Hal ini menjadi penting, terutama dalam menghadapi volatilitas pasar yang tinggi. Penelitian ini memfokuskan pada prediksi harga penutupan saham **MNC Bank**, sebuah bank swasta nasional yang aktif di Bursa Efek Indonesia. Pemilihan saham MNC Bank didasarkan pada karakteristik pergerakan harganya yang fluktuatif, serta perannya dalam sektor perbankan yang sangat dipengaruhi oleh dinamika ekonomi nasional. Dengan menggunakan algoritma KNN dan Linear Regression, penelitian ini bertujuan untuk mengukur sejauh mana akurasi masing-masing metode dalam memprediksi harga saham, serta memberikan kontribusi terhadap pengembangan sistem prediksi yang berbasis *machine learning*. Diharapkan hasil dari penelitian ini dapat menjadi acuan

bagi investor, pengembang sistem informasi keuangan, maupun peneliti yang tertarik dalam bidang analisis prediktif berbasis AI.

## **2. *BUSINESS UNDERSTANDING***

### **a) Permasalahan Dunia Nyata**

Dalam dunia investasi, terutama di pasar saham, kemampuan untuk memprediksi pergerakan harga saham secara akurat merupakan hal yang krusial. Namun pada kenyataannya, banyak investor baik individu maupun institusional menghadapi kesulitan dalam memproyeksikan harga saham karena tingginya volatilitas pasar dan banyaknya faktor eksternal yang memengaruhi. Hal ini semakin kompleks ketika menyangkut saham dari sektor perbankan seperti MNC Bank (BBMI), yang rentan terhadap dinamika ekonomi makro, isu keuangan global, dan kebijakan moneter domestik. Fluktuasi yang tinggi membuat analisis manual menjadi kurang andal, sehingga dibutuhkan pendekatan yang lebih sistematis dan berbasis data.

### **b) Tujuan utama dari proyek ini adalah :**

- Menerapkan algoritma regresi dalam memprediksi harga penutupan saham.
- Membandingkan performa dua algoritma regresi yang berbeda, yaitu K-Nearest Neighbor (KNN) dan Linear Regression, dalam konteks prediksi harga saham.
- Menganalisis sejauh mana tingkat akurasi dari kedua model dalam memberikan hasil prediksi terhadap data saham MNC Bank.
- Memberikan gambaran penerapan konsep kecerdasan buatan dalam dunia nyata, khususnya dalam sektor keuangan.

### **c) User/pengguna system**

Model prediksi harga saham yang dikembangkan dalam proyek ini ditujukan untuk berbagai kalangan pengguna, antara lain:

- Investor individu, yang membutuhkan informasi prediktif untuk merencanakan aksi beli atau jual saham.
- Analis keuangan, yang memerlukan model tambahan untuk memperkuat analisis teknikal mereka.
- Pengembang sistem rekomendasi investasi, yang dapat mengintegrasikan model ini ke dalam aplikasi trading atau sistem penasihat keuangan.
- Mahasiswa dan peneliti, khususnya di bidang ekonomi, keuangan, data science, atau kecerdasan buatan, sebagai bahan studi atau penelitian lanjutan.

#### **d) Manfaat implementasi AI**

Penerapan algoritma *machine learning* dalam prediksi harga saham memberikan sejumlah manfaat signifikan, di antaranya:

- Membantu pengambilan keputusan investasi yang lebih objektif, karena model bekerja berdasarkan data historis dan logika statistik, bukan spekulasi.
- Memberikan prediksi secara cepat dan otomatis, tanpa perlu analisis manual yang memakan waktu.
- Mengurangi risiko kesalahan dalam prediksi manual, yang seringkali dipengaruhi oleh bias atau keterbatasan pengetahuan investor.
- Meningkatkan efisiensi analisis saham, terutama ketika harus memantau banyak saham secara bersamaan.
- Dengan adanya sistem prediksi berbasis AI ini, diharapkan proses pengambilan keputusan di pasar saham menjadi lebih terukur, akurat, dan berbasis data.

### **3. DATA UNDERSTANDING**

#### **a) Dataset yang digunakan**

Dataset yang digunakan dalam proyek ini berasal dari file historis saham MNC Bank (Kaggle) yang dikumpulkan dalam format CSV. Data ini berisi informasi harga saham harian dan bersumber dari Kaggle. Dataset ini berisi 1.857 baris dan 13 kolom, dengan cakupan periode data mulai dari Januari 2014 hingga Agustus 2022, yang kemudian diolah untuk keperluan prediksi menggunakan machine learning. Dataset ini digunakan sebagai representasi saham individu untuk mempelajari pergerakan harga berdasarkan data historis dengan membandingkan 2 metode yaitu KNN dan Linier Regression.

#### **b) Deskripsi setiap fitur (atribut) :**

Meskipun dataset ini mencakup berbagai informasi seperti harga tertinggi (High), harga terendah (Low), volume, nilai transaksi, dan beberapa fitur turunan lainnya, hanya tiga kolom yang digunakan dalam proses analisis, yaitu:

- Date: Tanggal perdagangan saham,
- Open: Harga pembukaan saham pada hari tersebut,
- Close: Harga penutupan saham pada hari tersebut.

Ketiga atribut tersebut dipilih karena merepresentasikan data utama yang relevan dalam proses prediksi harga saham, khususnya untuk mengukur perubahan dan tren pergerakan harga harian

berdasarkan waktu. Data ini kemudian digunakan sebagai input dan target pada pemodelan machine learning.

**c) Ukuran dan format data:**

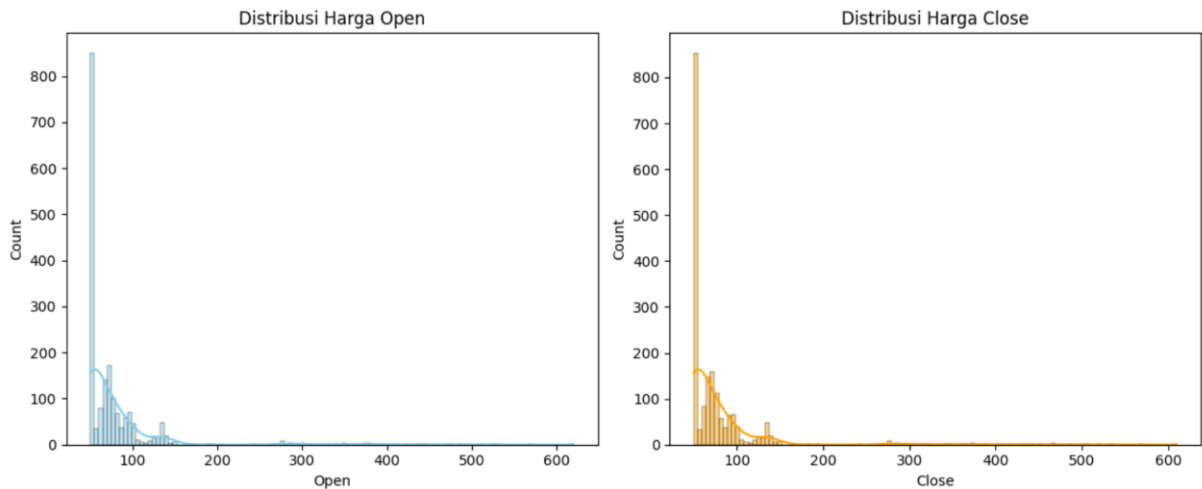
- Jumlah baris dan kolom: 1.857 baris dan 13 kolom.
- Format: CSV (.csv)

**d) Tipe data dan target *Forecasting*:**

- Tipe data: Numerik & Time-series. Dataset terdiri dari data numerik seperti harga saham dan open dan close, serta data bertipe waktu (Date) yang menunjukkan urutan kronologis perdagangan saham.
- Target: Regresi memprediksi nilai Close (harga penutupan). Nilai Close digunakan sebagai variabel target dalam pemodelan, dengan tujuan untuk memprediksi harga penutupan saham berdasarkan data historis.

#### **4. EXPLORATORY DATA ANALIS (EDA)**

**a) Visualisasi distribusi data histogram.**



Gambar 1. Grafik Histogram

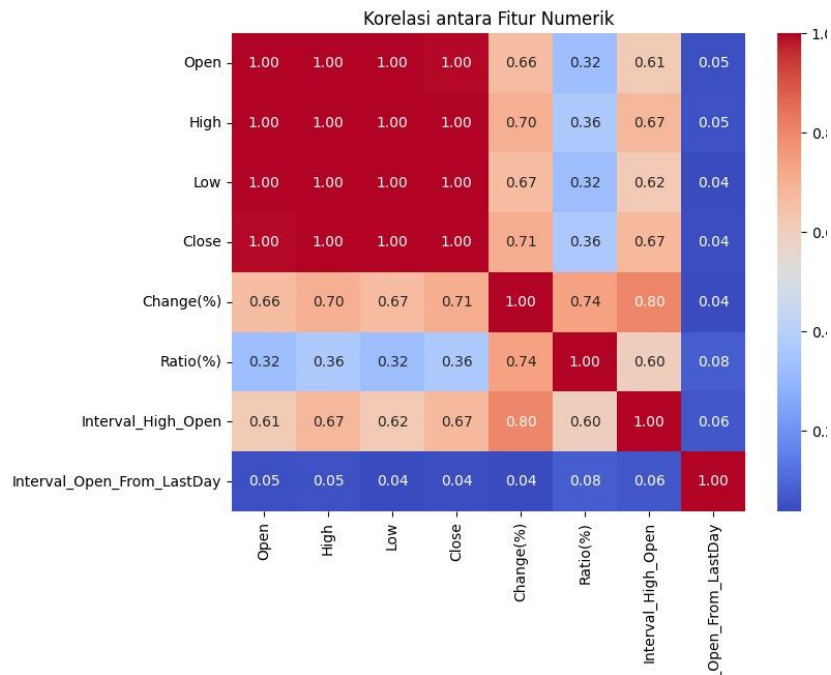
Gambar 1 menunjukkan distribusi harga pembukaan (Open) dan penutupan (Close) saham MNC Bank. Kedua variabel memiliki pola distribusi yang mirip, yaitu condong ke kanan (right-skewed), dengan mayoritas nilai terkonsentrasi di kisaran harga rendah sekitar Rp50–Rp100. Hanya sebagian kecil data yang menunjukkan harga di atas Rp100 hingga lebih dari Rp600. Pola ini mengindikasikan bahwa harga saham MNC Bank lebih sering berada di level rendah, dengan sesekali mengalami lonjakan. Distribusi seperti ini perlu diperhatikan karena dapat memengaruhi akurasi model prediksi, terutama pada algoritma regresi.



Gambar 2. Ten harga penutupan saham

Grafik ini menunjukkan saham yang berada dalam tren turun lama, kemudian melonjak ekstrem pada 2021 dengan volatilitas tinggi. Perubahan mendadak ini adalah fenomena penting yang perlu dicermati dalam investasi atau saat membangun model prediksi: model regresi standar yang mengasumsikan data stasioner kemungkinan besar tidak akan akurat menghadapi pola seperti ini.

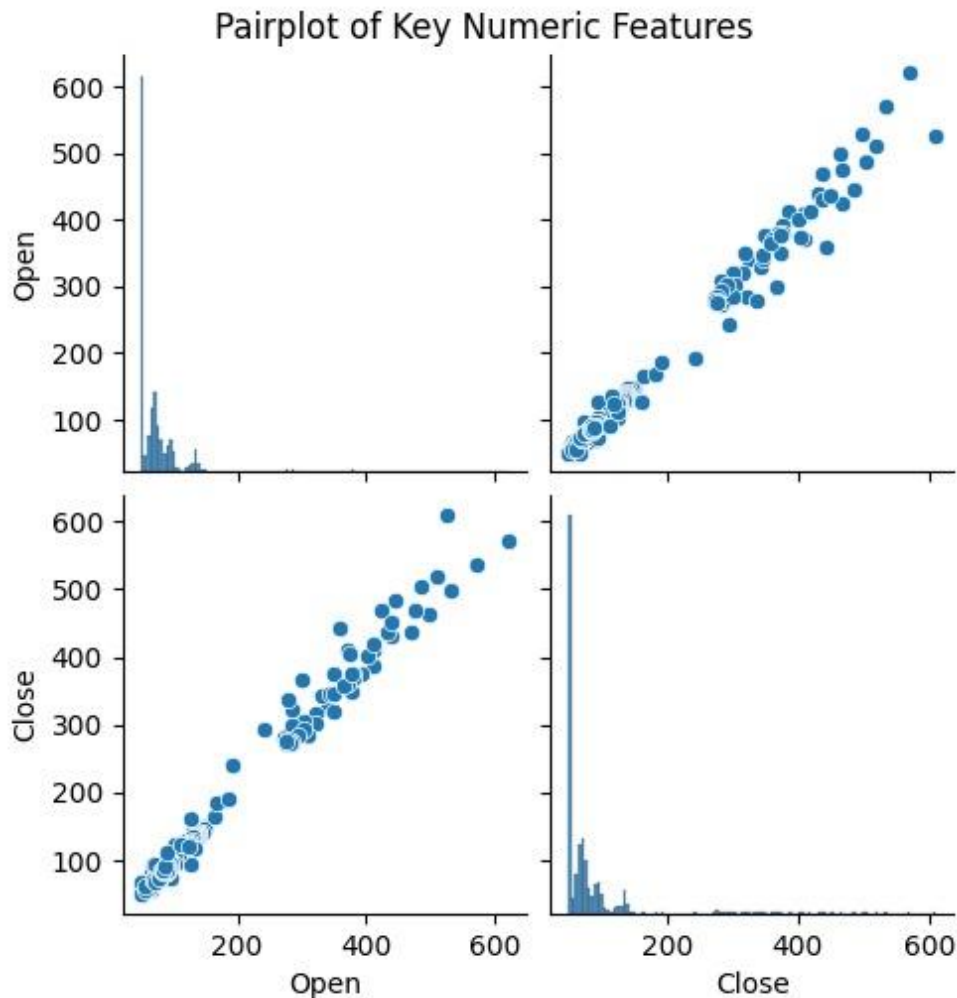
#### b) Analisis korelasi antar fitur dengan menggunakan heatmap dan Pairplot



Gambar 3. Heatmap Corelasi

Gambar 3 menunjukkan heatmap korelasi antar fitur numerik dalam dataset saham MNC Bank. Terlihat bahwa harga Open, High, Low, dan Close memiliki korelasi yang sangat tinggi

(nilai 1.00), menandakan hubungan yang sangat erat antar harga dalam satu hari perdagangan. Fitur turunan seperti `Change(%)` dan `Interval_High_Open` memiliki korelasi sedang terhadap harga, yang menunjukkan bahwa keduanya cukup relevan untuk analisis. Sementara itu, fitur `Interval_Open_From_LastDay` memiliki korelasi sangat lemah terhadap fitur lainnya, sehingga kurang berpengaruh dalam prediksi harga saham.



Gambar 4. Pairplot

Gambar 4 menunjukkan hubungan yang sangat kuat dan linear antara harga pembukaan (Open) dan penutupan (Close) saham MNC Bank. Titik-titik pada scatter plot membentuk garis diagonal, menandakan bahwa ketika harga Open naik, harga Close juga cenderung naik secara sebanding. Selain itu, histogram di diagonal menunjukkan bahwa mayoritas nilai berada di kisaran harga rendah. Visualisasi ini menguatkan bahwa fitur Open sangat relevan dalam memprediksi harga Close.

### c) Deteksi data tidak seimbang (imbalanced classes)

Pada dasarnya, dataset harga saham termasuk ke dalam jenis data regresi karena variabel target berupa nilai kontinu. Oleh karena itu, tidak terdapat kelas kategorikal yang memungkinkan dilakukannya analisis ketidakseimbangan kelas (*imbalanced classes*), seperti yang umum dilakukan pada permasalahan klasifikasi.

### d) Insight awal dari pola data

```
Statistik Deskriptif:
      Open      Close
count 1857.000000 1857.000000
mean   78.060312  78.022079
std    59.254470  59.800558
min    50.000000  50.000000
25%    50.000000  50.000000
50%    61.000000  61.000000
75%    80.000000  80.000000
max    620.000000  610.000000

Korelasi Open dan Close: 0.9955380886268013

Jumlah hari harga naik: 323
Jumlah hari harga turun: 1534
```

Gambar 5. Insight data

Data saham yang Anda miliki terdiri dari 1.857 hari perdagangan, mencakup periode yang panjang setara dengan sekitar tujuh hingga delapan tahun data harian. Dari statistik deskriptif yang diperoleh, harga rata-rata pembukaan (open) dan penutupan (close) masing-masing berada di kisaran 78. Angka ini menunjukkan bahwa secara umum, harga saham sering kali dibuka dan ditutup pada level yang hampir sama, tanpa pergeseran besar intraday yang konsisten ke satu arah.

Meskipun rata-ratanya rendah, standar deviasi harga yang mendekati 60 menunjukkan bahwa pergerakan harga cukup volatil relatif terhadap nilai rata-ratanya. Ini berarti harga sering mengalami fluktuasi yang signifikan, walaupun sebagian besar waktu tetap berada di level yang rendah. Hal ini juga tergambar dari nilai kuartil: 75% data memiliki harga penutupan di bawah 80, sementara nilai maksimum melonjak sangat tinggi hingga 610. Ini mengindikasikan adanya lonjakan harga yang ekstrem namun jarang terjadi, sehingga distribusi harga kemungkinan besar sangat miring ke kanan (skewed).

Nilai korelasi antara harga open dan close sangat tinggi, mendekati 0,9955. Ini menandakan adanya hubungan yang hampir linear sempurna: harga penutupan hariannya sangat erat mengikuti harga pembukaannya. Dengan kata lain, harga tidak banyak bergerak dari open ke close pada



sebagian besar hari, yang menunjukkan kecenderungan pasar yang cenderung stabil dalam intraday di luar periode lonjakan besar.

Dari sisi arah pergerakan harga harian, hanya terdapat 323 hari di mana harga penutupan lebih tinggi dari harga pembukaan, sedangkan 1.534 hari menunjukkan harga penutupan yang lebih rendah. Ini berarti sekitar 83% hari perdagangan berakhir dengan harga yang turun dibandingkan pembukaan, menunjukkan dominasi tren bearish selama periode data ini. Fakta ini sejalan dengan grafik tren harga yang Anda tampilkan sebelumnya, di mana harga cenderung turun dalam jangka panjang dengan hanya sedikit periode kenaikan tajam.

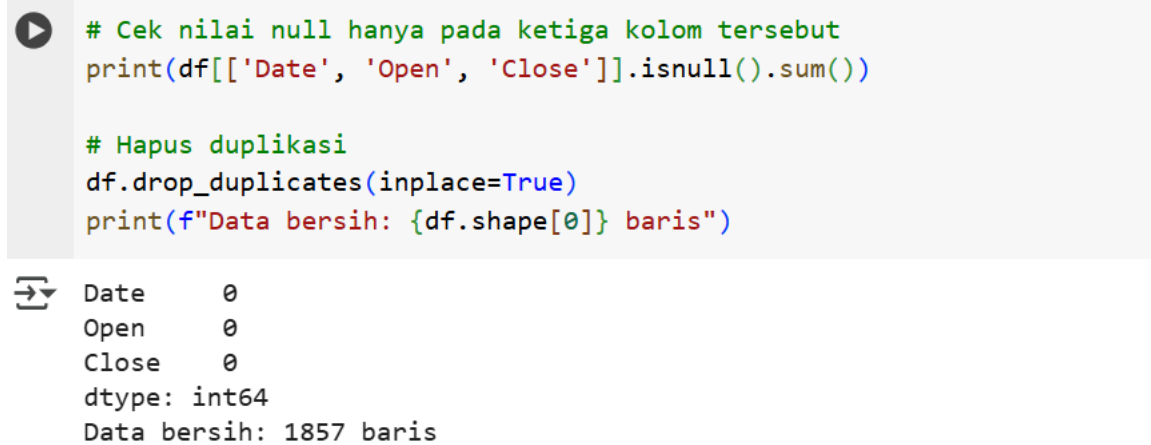
Secara keseluruhan, insight awal ini menunjukkan bahwa saham ini lebih sering mengalami penurunan harga dari hari ke hari, dengan volatilitas yang cukup besar meskipun sebagian besar harga tetap di kisaran rendah. Namun, lonjakan harga yang sangat tinggi pada momen tertentu membuat data menjadi tidak stasioner dan menimbulkan tantangan dalam membangun model prediksi yang mampu mengenali pola lonjakan tersebut. Analisis lebih mendalam, seperti melihat distribusi harga dan volatilitas per periode, akan sangat membantu dalam memahami dinamika saham ini secara lebih komprehensif.

## 5. EXPLORATORY DATA ANALIS (EDA)

### a) Pembersihan data (null value, duplikasi)

```
# Cek nilai null hanya pada ketiga kolom tersebut
print(df[['Date', 'Open', 'Close']].isnull().sum())

# Hapus duplikasi
df.drop_duplicates(inplace=True)
print(f>Data bersih: {df.shape[0]} baris")
```



```
Date      0
Open      0
Close     0
dtype: int64
Data bersih: 1857 baris
```

Gambar 6. Proses pembersihan data

Pada Gambar 6 dilakukan pemeriksaan nilai kosong (null) khusus pada tiga kolom utama yang akan digunakan dalam pemodelan, yaitu Date, Open, dan Close. Dari hasil pengecekan, ketiga kolom tersebut tidak memiliki nilai null sama sekali, yang berarti data pada kolom ini lengkap untuk seluruh 1.857 baris data.

Setelah itu, proses dilanjutkan dengan penghapusan duplikasi menggunakan `df.drop_duplicates()`. Hasilnya menunjukkan bahwa jumlah baris data tetap 1.857, menandakan bahwa dataset ini tidak mengandung baris yang persis sama secara keseluruhan.

Dengan demikian, data sudah dipastikan bersih dari nilai kosong maupun duplikasi, dan siap digunakan untuk tahapan pemrosesan dan modeling selanjutnya.

**b) Encoding data kategorik (label encoding, one-hot)**

Dalam penelitian ini, data yang digunakan terdiri dari kolom date (dalam format tanggal) dan dua fitur numerik, yaitu open dan close. Karena tidak terdapat fitur kategorik dalam dataset, maka proses encoding terhadap variabel kategorikal tidak diperlukan.

**c) Normalisasi data**

```
[14] #Normalisasi
      scaler = StandardScaler()
      X_train_scaled = scaler.fit_transform(X_train)
      X_test_scaled = scaler.transform(X_test)
```

Gambar 7. Normalisasi

Normalisasi dilakukan menggunakan StandardScaler untuk mengubah data fitur ke skala dengan rata-rata 0 dan standar deviasi 1. Data pelatihan di-*fit\_transform* untuk menghitung parameter skala, sedangkan data pengujian hanya di-*transform* dengan parameter dari train, sehingga mencegah data leakage. Normalisasi ini penting agar model dapat belajar lebih efektif tanpa bias akibat perbedaan skala antar fitur.

**d) Split data (train-test)**

```
[13] # Split data
      X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 8. Split data

Data dibagi menjadi data pelatihan dan pengujian dengan fungsi `train_test_split`, menggunakan proporsi 80% untuk train dan 20% untuk test (`test_size=0.2`). Pemisahan ini memastikan model dapat dilatih pada mayoritas data, lalu diuji pada data yang tidak pernah dilihat sebelumnya, sehingga performa model dapat dievaluasi secara objektif. Parameter `random_state=42` digunakan untuk memastikan hasil split konsisten dan dapat direproduksi.

## 6. MODELING

**a) Alasan Pemilihan algoritma Linear regression dan KNN**

Pemilihan algoritma Linear Regression dan K-Nearest Neighbor (KNN) dilakukan karena keduanya sesuai untuk tugas prediksi nilai numerik, yaitu harga penutupan saham.

Linear Regression dipilih karena merupakan metode yang sederhana, cepat, dan mudah diinterpretasikan. Model ini cocok digunakan untuk mengidentifikasi hubungan linier antara harga pembukaan dan harga penutupan saham, serta sering dijadikan baseline dalam prediksi berbasis data historis. Penelitian sebelumnya oleh tim dari ITN menunjukkan bahwa regresi linier efektif dalam memprediksi harga saham Bank BCA dengan akurasi yang tinggi ketika

ditopang oleh data historis yang lengkap serta tahapan praproses data yang tepat (Zapar et al., 2024).

Sementara itu, KNN Regression dipilih karena tidak bergantung pada asumsi linier. KNN dapat menangkap pola yang lebih kompleks dalam data, sehingga lebih fleksibel dalam menghadapi fluktuasi pasar saham yang tidak teratur.

Algoritma lain seperti Naive Bayes, SVM, atau Decision Tree tidak diprioritaskan karena kurang sesuai untuk regresi sederhana berbasis satu fitur numerik, atau memerlukan parameter tuning yang lebih kompleks. Oleh karena itu, fokus penelitian diarahkan pada dua model regresi yang paling relevan dan efisien.

#### b) Implementasi model (dengan kode)

```
#Model KNN
knn = KNeighborsRegressor(n_neighbors=5)
knn.fit(X_train_scaled, y_train)
knn_pred = knn.predict(X_test_scaled)
```

```
[16] #Model Regresi Linear
lr = LinearRegression()
lr.fit(X_train_scaled, y_train)
lr_pred = lr.predict(X_test_scaled)
```

Gambar 9. Model

Dua model regresi digunakan untuk memprediksi harga penutupan saham berdasarkan harga pembukaan yang telah dinormalisasi. Model pertama adalah *K-Nearest Neighbors Regressor* (KNN) dengan parameter 5 tetangga terdekat (`n_neighbors=5`), yang bekerja dengan memprediksi nilai berdasarkan rata-rata target dari data tetangga terdekat di ruang fitur.

Model kedua adalah *Linear Regression*, yang mempelajari hubungan linier antara fitur dan target. Kedua model dilatih menggunakan data pelatihan yang telah disiapkan.

Kemudian, kedua model digunakan untuk menghasilkan prediksi pada data pengujian. Langkah ini memungkinkan perbandingan performa kedua algoritma dalam memodelkan pola pergerakan harga saham.

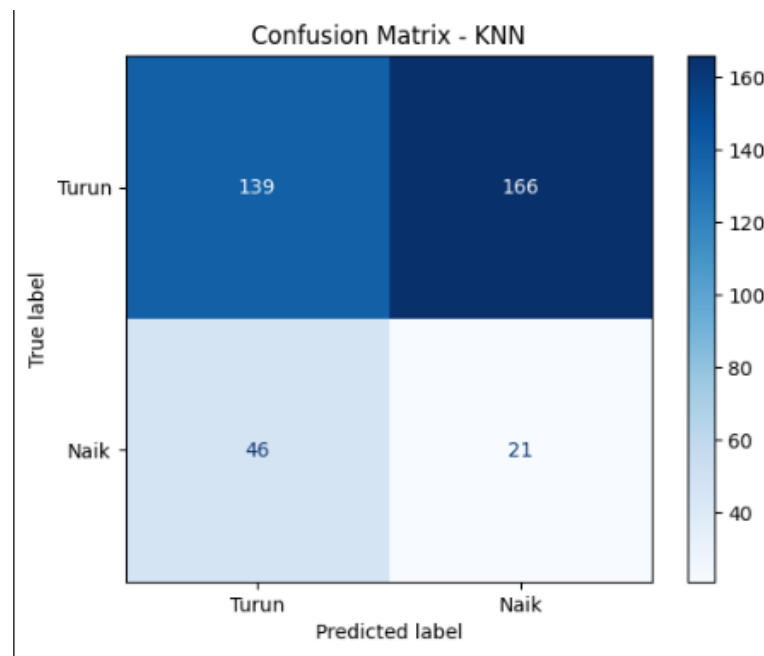
## 7. EVALUATION

Model dievaluasi menggunakan confusion matrix untuk mengetahui seberapa baik model dalam mengklasifikasikan arah pergerakan harga saham, yaitu antara kondisi turun (0) dan naik (1). Berikut ini adalah interpretasi dari dua confusion matrix yang dihasilkan:

### a) Confusion Matriks KNN

Visualisasi confusion matrix dari model K-Nearest Neighbors menunjukkan bahwa:

- Sebanyak 139 data diklasifikasikan benar sebagai *turun* (true negative).
- 21 data berhasil diklasifikasikan benar sebagai *naik* (true positive).
- Namun, terdapat 166 kesalahan prediksi naik padahal sebenarnya turun (false positive).
- Dan 46 kesalahan prediksi turun padahal sebenarnya naik (false negative).



Gambar 10. Confusion KNN

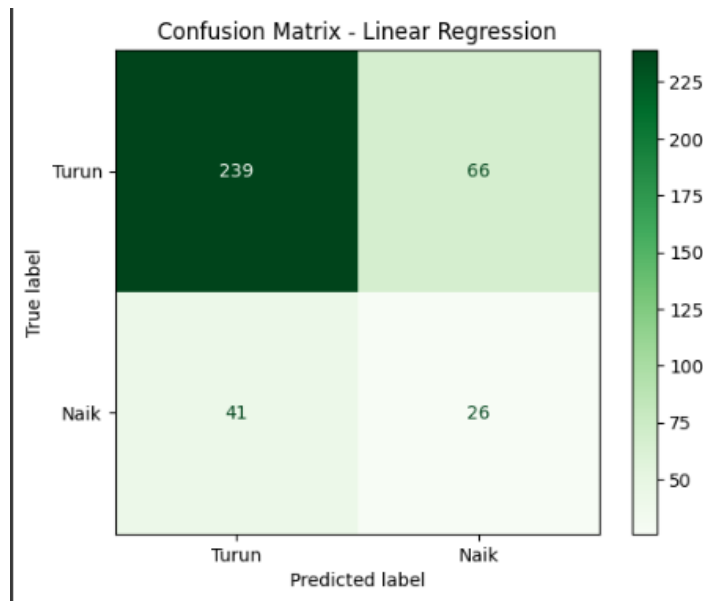
Artinya, model KNN memiliki kecenderungan untuk terlalu banyak memprediksi kelas “naik”, sehingga jumlah kesalahan prediksi menjadi cukup tinggi, terutama dalam membedakan pergerakan sebenarnya.

### b) Confusion Matriks Linear Regression (Konversi Klasifikasi)

Linear Regression pada dasarnya adalah model regresi, namun hasil prediksinya dikonversi ke dalam bentuk label klasifikasi (turun/naik), lalu dievaluasi menggunakan confusion matrix.

Hasilnya menunjukkan:

- 239 data benar diklasifikasikan sebagai turun, dan 26 data benar diklasifikasikan sebagai naik.
- Terdapat 66 data salah diklasifikasikan naik, padahal sebenarnya turun.
- Serta 41 data salah diklasifikasikan turun, padahal sebenarnya naik.



Gambar 11. KNN LR

Dibandingkan KNN, Linear Regression menunjukkan jumlah prediksi benar yang lebih tinggi, serta jumlah kesalahan (false positive dan false negative) yang lebih rendah, meskipun model ini awalnya tidak dirancang untuk klasifikasi.

### c) Metrik evaluasi

Evaluasi performa dilakukan pada dua model yaitu K-Nearest Neighbors (KNN) dan Linear Regression (yang telah dikonversi ke label klasifikasi). Hasil evaluasi ditampilkan melalui metrik akurasi, precision, recall, dan f1-score.

Hasil evaluasi model KNN dan LR:

```

=== KNN ===
Akurasi   : 0.4301
Precision : 0.1123
Recall    : 0.3134
F1-score  : 0.1654

=== Linear Regression ===
Akurasi   : 0.7124
Precision : 0.2826
Recall    : 0.3881
F1-score  : 0.3270

```

Gambar 12. Metriks evaluasi

Model KNN menunjukkan performa yang kurang optimal pada data ini. Meskipun recall cukup tinggi, precision dan f1-score masih rendah, yang mengindikasikan bahwa model sering salah dalam memprediksi kelas naik/turun secara akurat.

Model Linear Regression yang awalnya digunakan untuk prediksi numerik justru memberikan hasil klasifikasi yang lebih baik dibanding KNN, baik dari sisi akurasi maupun keseimbangan antara precision dan recall.

#### **d) Kinerja Model**

Model KNN menunjukkan performa yang kurang baik, dengan akurasi hanya sekitar 43.01%. Nilai precision yang sangat rendah (11.23%) mengindikasikan bahwa sebagian besar prediksi yang diklaim sebagai “naik” ternyata salah. Selain itu, recall yang hanya 31.34% menunjukkan bahwa model gagal mendeteksi banyak instance yang sebenarnya merupakan kelas “naik”. Nilai f1-score yang rendah (16.54%) menandakan bahwa model tidak seimbang dalam hal ketepatan dan cakupan prediksi. Secara keseluruhan, KNN tidak cukup mampu menangkap pola yang dibutuhkan untuk klasifikasi arah pergerakan harga saham, mungkin karena fitur-fitur input tidak cukup kuat untuk model berbasis jarak.

Sebaliknya, Linear Regression (yang seharusnya digunakan untuk regresi, namun dikonversi ke klasifikasi) menunjukkan kinerja yang lebih baik. Model ini menghasilkan akurasi sebesar 71.24%, dengan precision 28.26%, recall 38.81%, dan f1-score 32.70%. Walaupun precision dan recall masih tergolong rendah secara absolut, hasilnya jauh lebih baik dibandingkan dengan KNN. Hal ini menunjukkan bahwa Linear Regression lebih mampu mengenali pola tren harga saham secara umum, meskipun tidak secara spesifik mengklasifikasikan semua instance dengan sempurna.

Dengan melihat metrik-metrik tersebut, dapat disimpulkan bahwa Linear Regression lebih layak digunakan dalam kasus ini dibandingkan KNN. Selain itu, hasil evaluasi ini juga menggambarkan bahwa prediksi arah pergerakan harga saham bukanlah masalah klasifikasi yang sederhana dan memerlukan pendekatan yang lebih kompleks atau fitur tambahan yang lebih informatif.

### **8. KESIMPULAN & REKOMENDASI**

Proyek ini membandingkan dua algoritma regresi, yaitu K-Nearest Neighbors (KNN) dan Linear Regression, untuk memprediksi harga penutupan saham MNC Bank berdasarkan data historis. Berdasarkan hasil evaluasi performa model, diketahui bahwa Linear Regression memberikan hasil klasifikasi yang lebih baik, dengan akurasi sebesar 71.24%, precision 28.26%, recall 38.81%, dan f1-score 32.70%. Regresi linier menunjukkan performa yang cukup baik dalam memprediksi harga saham, namun model seperti Neural Network mampu

menangkap pola non-linear yang lebih kompleks, terutama pada dataset saham jangka panjang(Cahyono, 2025).

Secara umum, tujuan proyek dinyatakan tercapai, yakni:

- Mengimplementasikan dua algoritma regresi untuk prediksi harga saham.
- Melakukan evaluasi performa terhadap kedua model.
- Menunjukkan bahwa metode machine learning dapat digunakan untuk membantu menganalisis pergerakan harga saham secara data-driven.

Meskipun akurasi model belum maksimal, terutama pada KNN, hasil Linear Regression telah membuktikan bahwa pendekatan berbasis AI dapat diaplikasikan dalam konteks nyata.

**Kelebihan:**

- Linear Regression sederhana, cepat, dan cocok untuk data dengan pola linier seperti hubungan harga Open dan Close saham.
- KNN bersifat non-parametrik, sehingga dapat menangkap pola non-linier, meskipun dalam eksperimen ini hasilnya belum optimal.

**Keterbatasan:**

- Model KNN cenderung terlalu sensitif terhadap noise dan rentan terhadap overfitting jika data tidak bersih atau tidak relevan.
- Dataset yang digunakan hanya menggunakan sebagian kecil fitur (Open, Close), sehingga tidak cukup kaya untuk menangkap keseluruhan pola pasar.
- Linear Regression sebagai metode regresi tidak secara native dirancang untuk klasifikasi, sehingga perlu konversi label yang menyederhanakan kompleksitas pergerakan pasar.

Agar sistem prediksi harga saham dapat dikembangkan lebih lanjut dan memberikan hasil yang lebih akurat, beberapa langkah yang disarankan antara lain:

1. Menggunakan dataset yang lebih besar dan lebih variatif, misalnya dengan mencakup indikator teknikal lain seperti moving average, RSI, MACD, dan lainnya.
2. Mengeksplorasi algoritma yang lebih canggih, seperti Random Forest, Gradient Boosting (XGBoost), atau Long Short-Term Memory (LSTM) untuk menangkap pola jangka panjang dan temporal.
3. Melakukan hyperparameter tuning pada KNN dan model lain untuk mencari konfigurasi terbaik yang dapat meningkatkan performa prediksi.
4. Melibatkan lebih banyak fitur dari data pasar saham, seperti volume, perubahan harian (daily change), dan indikator sentimen pasar.

5. Menggunakan teknik validasi silang (cross-validation) untuk mengevaluasi kinerja model secara lebih menyeluruh dan menghindari hasil yang terlalu tergantung pada satu pembagian data.

## 9. DAFTAR PUSTAKA

- Eko Waluyo, D., Wikan Kinasih, H., Paramita, C., Pergiwati, D., Adi Rafrastara, F., Nohan, R., Studi Akuntansi, P., Ekonomi dan Bisnis, F., & Dian Nuswantoro, U. (n.d.). Implementasi Algoritma Regresi pada Machine Learning untuk Prediksi Indeks Harga Saham Gabungan. *Universitas Dian Nuswantoro, Semarang Jln. Imam Bonjol*, 9(1).
- Purnama Sari, E., Bachri, S. M., Atnang, M., Fajar, N., Studi Teknologi Informasi, P., Sains Teknologi dan Kesehatan, F., Sains Teknologi dan Kesehatan, I., & Kendari, A. (2024). Studi Literatur Deep Learning dan Machine Learning untuk Analisis dan Prediksi Pasar Saham: Metodologi, Representasi Data dan Studi Kasus. In *Jurnal Teknologi dan Sains Modern* (Vol. 1, Issue 1). <https://journal.scitechgrup.com/index.php/jtsm>
- Triya, P., Suarna, N., & Nuris, N. D. (2024). PENERAPAN MACHINE LEARNING DALAM MELAKUKAN PREDIKSI HARGA SAHAM PT. BANK MANDIRI (PERSERO) TBK DENGAN ALGORITMA LINEAR REGRESSION. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Issue 1).
- Cahyono, W. (2025). *Analisis Perbandingan Algoritma Regresi Linier dengan Neural Network untuk Prediksi Harga Saham Comparative Analysis of Linear Regression and Neural Network Algorithms*. 14, 1879–1896.
- Zapar, R., Pratama, D., Kaslani, K., Rohmat, C. L., & Faturrohman, F. (2024). Penerapan Model Regresi Linier Untuk Prediksi Harga Saham Bank Bca Pada Bursa Efek Indonesia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 196–202. <https://doi.org/10.36040/jati.v8i1.8215>